

GenCore version 5.1.6
copyright (c) 1993 - 2003 Compugen Ltd.

OM protein - protein search, using sw model

Run on : May 29, 2003, 15:21:44 ; Search time 22 Seconds
(without alignments)
152.708 Million cell updates/sec

Title: US-09-924-102-2

Perfect score: 81

Sequence: I MLLSTHLFLFLFLFLSYSL.....RGGQQGRRGGTADTGGMFLS 81

Scoring table: OLIGO
Gapop 60.0 , Gapext 60.0

Searched: 112892 seqs, 41476328 residues

Word size : 0

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Listing first 45 summaries

Database : Swissprot_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match Length	DB ID	Description
1	8	9.9	65	RPR_DRDOME
2	8	9.9	284	1 SUHA_HUMAN
3	8	9.9	284	1 SUHA_MACFA
4	8	9.9	587	1 ALU2_HUMAN
5	8	9.9	591	1 ALU8_HUMAN
6	8	9.9	116	1 FLAD_XENFE
7	8	8.6	285	1 INVO_CANFE
8	7	8.6	304	1 Y687_HAEGIN
9	7	8.6	337	1 YK68_YEAST
10	7	8.6	352	1 CITEC_ECOLI
11	7	8.6	368	1 ROXI YEAST
12	7	8.6	591	1 ALUL_HUMAN
13	7	8.6	603	1 ALU4_HUMAN
14	7	8.6	804	1 SYL_BACSU
15	7	8.6	805	1 SYL_BACCH
16	7	8.6	889	1 C122_HUMAN
17	7	8.6	1230	1 UGS4_SOLTU
18	7	8.6	1 HSP1_ISOMA	
19	6	7.4	65	1 HSP1_ISOMA
20	6	7.4	68	1 HSP1_PERGU
21	6	7.4	93	1 HIS2_NYCTU
22	6	7.4	118	1 RS20_SCHPO
23	6	7.4	134	1 SIZL_HUMAN
24	6	7.4	140	1 RL23_DRDOME
25	6	7.4	153	1 IF1A_YEAST
26	6	7.4	154	1 YK01_AERPE
27	6	7.4	160	1 HBLL_ARATH
28	6	7.4	168	1 RR7_CHLRE
29	6	7.4	174	1 ILIX_BOVIN
30	6	7.4	176	1 DPSA_SYNPT
31	6	7.4	189	1 COAT_TMV
32	6	7.4	189	1 COAT_TIMVA
193	6	7.4	1 KITH_HAEIN	

ALIGNMENTS

34	6	7.4	201	1 COAE_BRACHD
35	6	7.4	206	1 COX3_BACFT
36	6	7.4	211	1 CUDD_MOUSE
37	6	7.4	214	1 GRP2_NICSY
38	6	7.4	231	1 TRMD_MCGE
39	6	7.4	252	1 THI4_PYRAB
40	6	7.4	255	1 TRPA_AQUAE
41	6	7.4	262	1 RP32_HAEGIN
42	6	7.4	281	1 RP32_CITFR
43	6	7.4	284	1 RP32_ECOLI
44	6	7.4	284	1 RP32_ECOLI
45	6	7.4	285	1 RP32_ECOLI

09K657	bacillus ha
004442	bacillus fi
02z04	muis musculu
027484	nictotiana s
P47683	mycoplasma
09v018	pyrococcus
059082	pyrococcus
067502	quifex aeo
P4404	hemophilus
P11539	citrobacter
P00580	escherichia
P50508	enterobacte

RA Wang Z.-Y., Wasserman D.A., Weinstock G.M., Weissbach J., Weissbach H., Hume R.;
 RA Williams S.M., Woodcock T., Worley K.C., Wu D., Yang S., Yao Q.A.,
 RA Ye J., Yeh R.-F., Zaveri J.S., Zhan M., Zhang G., Zhao O., Zheng L.,
 RA Zheng X.H., Zhong F.N., Zhong W., Zhou X., Zhu S., Zhu X., Smith H.O.,
 RA Gibbs R.A., Myers E.W., Rubin G.M., Venter J.C.,
 RA "The genome sequence of *Drosophila melanogaster*";
 RL *Science* 287:2185-2195(2000).
 CC -I- FUNCTION: PLAYS A CENTRAL AND GLOBAL REGULATORY FUNCTION FOR THE
 CC INITIATION OF APOPTOSIS. ECTOPIC EXPRESSION IN THE DEVELOPING EYE
 CC RESULTS IN A SMALL EYE OWING TO EXCESS CELL DEATH.
 CC -I- DEVELOPMENTAL STAGE: EXPRESSION CORRESPONDS TO THE PATTERN OF
 CC PROGRAMMED CELL DEATH IN THE EMBRYO.
 CC -I- SIMILARITY: LIMITED AT THE N-TERMINAL, TO HID AND GRIM.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC EMBL; U11631; AAA18983.1; -.
 DR EMBL; AB003520; AAF49264.1; -.
 DR FBgn001706; rpr.
 KW Apoptosis.
 SEQUENCE 35 QKEQQLR 42
 Query Match 9.9%; Score 8; DB 1; Length 65;
 Best Local Similarity 100.0%; Pred. No. 0.17; 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 SQ RN
 Db 19 QKEQQLR 26
 QY 35 QKEQQLR 42
 DR 65 AA; 7682 MW; 57F231379AEEA3C CRC64;
 DR Kong A.-N.T., Yang L., Ma M., Tao D., Bjornsson T.D.,
 DR "Molecular cloning of the alcohol/hydroxysteroid form (hsta) of
 DR sulfotransferase from human liver";
 DR *Biochem. Biophys. Res. Commun.* 187:448-454(1992).
 RN [7]
 RP SEQUENCE FROM N.A.
 RC TISSUE=liv;
 RA MEDLINE:9239364; PubMed=1520333;
 RA Strauberg R.;
 RL Kong A.-N.T., Yang L., Ma M., Tao D., Bjornsson T.D.,
 RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
 CC -I- FUNCTION: CATALYZES THE SULFATION OF STEROIDS AND BILE ACIDS IN
 CC THE LIVER AND ADRENAL GLANDS.
 CC -I- CATALYTIC ACTIVITY: 3'-phosphoadenyl sulfate + an alcohol =
 CC adenosine 3',5'-bisphosphate + an alkyl sulfate.
 CC -I- SUBUNIT: HOMODIMER.
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 01-JUN-2002 (Rel. 41, Last annotation update)
 DE Alcohol sulfotransferase (EC 2.8.2.2) (Hydroxysteroid
 DE Sulfotransferase) (HST) (Dehydroepiandrosterone sulfotransferase)
 DE (DHSA-ST) (S92) (ST2A3).
 GN SULT2A1 OR S92 OR HST.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Homindae; Homo.
 OX NCBI_TAXID=9606;
 RN [1]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 60-64; 104-119 AND 273-284.
 TISSUE=liver;
 RX MEDLINE:93143674; PubMed=7678732;
 RA Comer K.A., Falany J.L., Falany C.N.;
 RT "Cloning and expression of human liver dehydroepiandrosterone
 RT sulfotransferase";
 RL Blochem. J. 289:233-240(1993).
 RN [2]
 RP SEQUENCE FROM N.A., AND SEQUENCE OF 80-107 AND 176-198.
 RC TISSUE=liver;
 RX MEDLINE:92269778; PubMed=1588921;
 RA Otterness D.M., Weben E.D., Wood T.C., Watson R.W.G., Madden B.J.,
 RA McCormick D.J., Weinshilboum R.M.;
 RT "Human liver dehydroepiandrosterone sulfotransferase: molecular
 RT cloning and expression of cDNA";
 RL Mol. Pharmacol. 41:865-872(1992).
 RN [3]
 RP SEQUENCE FROM N.A.
 RC TISSUE=adrenal gland;
 RX MEDLINE:96034512; PubMed=7589785;

RA Forbes K.J., Hagen M., Coughtrie M.W.H., Glatt H.R., Hume R.;
 RA "Human fetal adrenal hydroxysteroid sulphotransferase: cDNA cloning,
 RT stable expression in V79 cells and functional characterisation of the
 RT expressed enzyme.>";
 RT *Mol. Cell. Endocrinol.* 112:53-60(1995).
 RL [4]
 RN SEQUENCE FROM N.A.
 RP MEDLINE:9532029; PubMed=7598806;
 RX LAU-The V., Duport I., Paquet N., Reinmitz G., Labrie F.;
 RA "Structural characterization and expression of the human
 RA Weinshilboum R.M.;"
 RT "Human dehydroepiandrosterone sulfotransferase gene: molecular
 RT cloning and structural characterization.>";
 RL *DNA Cell Biol.* 14:511-518(1995).
 RN [6]
 RP SEQUENCE FROM N.A.
 RC TISSUE=liver;
 RA MEDLINE:9239364; PubMed=1520333;
 RA Strauberg R.;
 RL Kong A.-N.T., Yang L., Ma M., Tao D., Bjornsson T.D.,
 RL Submitted (JAN-2002) to the EMBL/GenBank/DDBJ databases.
 CC -I- FUNCTION: CATALYZES THE SULFATION OF STEROIDS AND BILE ACIDS IN
 CC THE LIVER AND ADRENAL GLANDS.
 CC -I- CATALYTIC ACTIVITY: 3'-phosphoadenyl sulfate + an alcohol =
 CC adenosine 3',5'-bisphosphate + an alkyl sulfate.
 CC -I- SUBUNIT: HOMODIMER.
 CC -I- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -I- TISSUE SPECIFICITY: LIVER, ADRENAL AND AT LOWER LEVEL IN THE ADRENAL
 CC THAN THE LIVER AND THE KIDNEY.
 CC -I- PTM: THE N-TERMINUS IS BLOCKED.
 CC -I- MISCELLANEOUS: ESTROGENS PRESENT IN MATERNAL CIRCULATION IS
 CC REDOMINANTLY DERIVED FROM FETAL DEHYDROEPANDROSTERONE SULFATE
 CC WHICH IS HYDROLYZED TO ESTROGENS IN PLACENTA.
 CC -I- SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 DR EMBL; I20000; AAA35758.1; -.
 DR EMBL; X70222; CAQ49755.1; -.
 DR EMBL; U08024; AAC17749.1; -.
 DR EMBL; U08025; AAC17750.1; -.
 DR EMBL; X84816; CAQ59274.1; -.
 DR EMBL; L36196; AAC75491.1; -.
 DR EMBL; L36191; AAC75491.1; JOINED.
 DR EMBL; L36192; AAC75491.1; JOINED.
 DR EMBL; L36193; AAC75491.1; JOINED.
 DR EMBL; L36194; AAC75491.1; JOINED.
 DR EMBL; L36195; AAC75491.1; JOINED.
 DR EMBL; U13061; AAC51353.1; JOINED.
 DR EMBL; U13056; AAC51353.1; JOINED.
 DR EMBL; U13057; AAC51353.1; JOINED.
 DR EMBL; U13058; AAC51353.1; JOINED.
 DR EMBL; U13059; AAC51353.1; JOINED.
 DR EMBL; U13060; AAC51353.1; JOINED.

Best Local Similarity 100.0%; Pred. No. 0.59; 0; Indels 0; Gaps 0;

Matches	8; Conservative 0; Mismatches 0;
Qy	48 FRSETLRK 55
Db	17 FRSETLRK 24

Query Match 9.9%; score 8; DB 1; Length 284;
 Best Local Similarity 100.0%; Pred. No. 0.59; 0; Mismatches 0; Indels 0; Gaps 0;

Matches 8; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 48 FRSETLRK 55

Db 17 FRSETLRK 24

RESULT 3

SUHA_MACFA	STANDARD;	PRT;	284 AA.
ID SUHA_MACFA			
AC P52842;			
DT 01-OCT-1996 (Rel. 34, Created)			
DT 01-OCT-1996 (Rel. 34, Last sequence update)			
DT 16-OCT-2001 (Rel. 40, Last annotation update)			
DE Alcohol sulfotransferase (EC 2.8.2.2) (Hydroxysteroid sulfotransferase) (HST).			
DE SULTA1 OR STD.			
OS Macaca fascicularis (Crab eating macaque) (<i>Cynomolgus monkey</i>).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Cercopithecidae;			
OX NCBI_TaxID=9541;			
RN [1]			
SEQUENCE FROM N.A.			
RP TISSUE-LIVER;			
RA Ogura K., Satsukawa M., Kato K., Okuda H., Watabe T.;			
RL Submitted (MAY-1996) to the EMBL/GenBank/DDBJ databases.			
CC - FUNCTION: CATALYSES THE SULFATION OF STEROIDS AND BILE ACIDS IN THE LIVER AND ADRENAL GLANDS (BY SIMILARITY).			
CC - CATALYTIC ACTIVITY: 3'-phosphoadenylylsulfate + an alcohol = adenosine 3',5'-bisphosphate + an alkyl sulfate.			
CC - SUBUNIT: HOMODIMER (BY SIMILARITY).			
CC - SUBCELLULAR LOCATION: CYTOPLASMIC (BY SIMILARITY).			
CC - SIMILARITY: BELONGS TO THE SULFOTRANSFERASE FAMILY.			
CC - This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@sb-sib.ch).			
CC - EMBL: D85521; BAA12823.1; - .			
CC - HSSP: P49891; IAUQ;			
CC - InterPro: IPR000853; Sulfotransferase.			
CC - Pfam: PF00685; Sulfotransferase_1.			
CC - DR: Probm; PDB001218; Sulfotransferase; 1.			
CC - DR: Transf erase; Steroid metabolism.			
CC - KW: BY SIMILARITY.			
CC - FT: BINDING 248 254 PAPS BINDING SITE (POTENTIAL).			
CC - SQ: SEQUENCE 284 AA; 33789 MW; 3C89C759/833EBR1 CRC64;			
CC - Query Match 9.9%; Score 8; DB 1; Length 284;			

RESULT 4

ALU2_HUMAN	STANDARD;	PRT;	587 AA.
ID ALU2_HUMAN			
AC P39189;			
DT 01-FEB-1995 (Rel. 31, Created)			
DT 01-FEB-1995 (Rel. 31, Last sequence update)			
DT 16-OCT-2001 (Rel. 40, Last annotation update)			
DE Alu subfamily SB sequence contamination warning entry.			
OS Homo sapiens (Human).			
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;			
OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.			
OX NCBI_TaxID=9606;			
RN [1]			
SEQUENCE FROM N.A.			
RP MEDLINE=950211758; PubMed=7935834;			
RA Claverie J.-M., Makalowski W.;			
RT "Alu alert.";			
RL Nature 371:752-752(1994).			
RP CONCEPT.			
RX MEDLINE=922411891; PubMed=1572661;			
RA Claverie J.-M.;			
RT "Identifying coding exons by similarity search: alu-derived and other potentially misleading protein sequences.";			
RT Genomics 12:838-841(1992).			
RN [3]			
RP ALU FAMILIES CLASSIFICATION.			
RX MEDLINE=88333009; PubMed=3138422;			
RA Quentin Y.;			
RT "The Alu family developed through successive waves of fixation closely connected with primate lineage history.";			
RL J. Mol. Evol. 27:194-202(1998).			
RN [4]			
RP ALU FAMILIES CLASSIFICATION.			
RX MEDLINE=9178815; PubMed=1706781;			
RA Jurka J., Milosavljevic A.;			
RT "Reconstruction and analysis of human Alu genes.";			
J. Mol. Evol. 32:105-121(1991).			
CC - MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.			
CC - MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.			
CC - CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO ACID SEQUENCES.			
CC - CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER, CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED cDNAs LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS, CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES BEING REPORTED.			
CC - CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A PART OF ALU-TRANSLATED ENTRY MAY HAVE BEEN ARTIFICIALLY INCLUDED IN THE			

CC - CODING NUCLEOTIDE SEQUENCE.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

DR EMBL; UI4569; ; NOT_ANNOTATED_CDS.

KW Hypothetical protein.

FT DOMAIN 1 96 FRAME-1.

FT DOMAIN 100 194 FRAME-2.

FT DOMAIN 198 292 FRAME-3.

FT DOMAIN 296 391 FRAME-4.

FT DOMAIN 395 489 FRAME-5.

FT DOMAIN 493 587 FRAME-6.

SQ SEQUENCE 587 AA; 63703 MW; 3EAB3D3E3929203 CRC64;

Query Match 9.9%; Score 8; DB 1; Length 587;

Best Local Similarity 100.0%; Pred. No. 1.1;

Matches 8; Conservative 0; Mismatches 0; Indels 0;

Gaps 0;

OY 19 SLGDRARL 26

Db 183 SLGDRARL 190

RESULT 5

ALU_HUMAN STANDARD; PRT; 587 AA.

ID ALU3_HUMAN STANDARD; PRT; 587 AA.

AC P39190;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Alu subfamily SBI sequence contamination warning entry.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606; [1]

RN SEQUENCE FROM N.A.; MEDLINE=95021758; PubMed=7935834;

RX Claverie J.-M.; Makalowski W.;

RT "Alu alert.";

RL Nature 371:752-752(1994).

RN [2]

RP CONCEPT; MEDLINE=92241891; PubMed=1572661;

RX Claverie J.-M.;

RT "Identifying coding exons by similarity search: alu-derived and other potentially misleading protein sequences.";

RL Genomics 12:838-841(1992).

RN [3]

RP ALU_FAMILIES CLASSIFICATION; MEDLINE=88333009; PubMed=3138422;

RX Quentin Y.;

RT "The Alu family developed through successive waves of fixation closely connected with primate lineage history.";

RL J. Mol. Evol. 27:194-202(1998).

RN [4]

RP ALU_FAMILIES CLASSIFICATION; MEDLINE=9117815; PubMed=1706781;

RX Jurka J.; MilosavJevic A.;

RT "Reconstruction and analysis of human Alu genes.";

RL J. Mol. Evol. 32:105-121(1991).

CC - MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU REPEATS.

CC - MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP

CC CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.

CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER

CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO

CC ACID SEQUENCES.

-1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPersed IN HUMAN AND

PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE

ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN

ALU-DERIVED SEQUENCES IN 5', OR 3', UNTRANSLATED REGIONS. HOWEVER,

CNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAs

LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU

ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A

GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,

CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE

OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A

CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING

DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH

THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES

BEING REPORTED.

-1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE

WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A

PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE

CODING NUCLEOTIDE SEQUENCE.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

CC DR EMBL; UI4569; ; NOT_ANNOTATED_CDS.

KW Hypothetical protein.

FT DOMAIN 1 96 FRAME-1.

FT DOMAIN 100 194 FRAME-2.

FT DOMAIN 198 292 FRAME-3.

FT DOMAIN 296 391 FRAME-4.

FT DOMAIN 395 489 FRAME-5.

FT DOMAIN 493 587 FRAME-6.

SQ SEQUENCE 587 AA; 63573 MW; 85C4155726DEF235 CRC64;

Query Match 9.9%; Score 8; DB 1; Length 587;

Best Local Similarity 100.0%; Pred. No. 1.1;

Matches 8; Conservative 0; Mismatches 0; Indels 0;

Gaps 0;

OY 19 SLGDRARL 26

Db 183 SLGDRARL 190

RESULT 6

ALU8_HUMAN STANDARD; PRT; 591 AA.

ID ALU8_HUMAN STANDARD; PRT; 591 AA.

AC P39195;

DT 01-FEB-1995 (Rel. 31, Created)

DT 01-FEB-1995 (Rel. 31, Last sequence update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE Alu subfamily S8X sequence contamination warning entry.

OS Homo sapiens (Human).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Buteleostomi;

OC Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

NCBI_TaxID=9606; [1]

RN SEQUENCE FROM N.A.; MEDLINE=7935834;

RX Claverie J.-M.; Makalowski W.;

RT "Alu alert.";

RL Nature 371:752-752(1994).

[2]

RP CONCEPT; MEDLINE=92241891; PubMed=1572661;

RX Claverie J.-M.;

*Identifying coding exons by similarity search: alu-derived and other

RT potentially misleading protein sequences.;"
 RL Genomics 12:838-841(1992).
 [3]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=88333009; PubMed=3138422;
 RA Quentin Y;
 RT "the Alu family developed through successive waves of fixation
 closely connected with primate lineage history.";
 RL Mol. Evol. 27:194-202(1988).
 [4]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=91178815; PubMed=1706781;
 RA Jurka J, Milosavljevic A.;
 RT "reconstruction and analysis of human Alu genes.";
 J. Mol. Evol. 32:105-121(1991).
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE
 THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
 CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
 FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
 REPEATS.
 CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
 CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.
 CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
 POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
 ACID SEQUENCES.
 CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPersed IN HUMAN AND
 PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
 ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
 ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR Rearranged cDNAs
 LIGHTED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
 ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
 GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
 CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
 OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
 CLOTHING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
 DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
 THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
 BEING REPORTED.
 CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
 WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
 PART OF ALU REPEAT MAY HAVE BEEN ARTIFACTUALLY INCLUDED IN THE
 CC CODING NUCLEOTIDE SEQUENCE.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).

CC -----

DR EMBL; AU12838; CAB41407.1; -
 KW Transcription; Transcription regulation; Sigma factor;
 CC DNA-directed RNA Polymerase; DNA-binding; Flagellin.
 CC SQ SEQUENCE 116 AA; 1303 MW; 89DBP#817532828 CRC64;

Query Match Best Local Similarity 8.6%; Score 7; DB 1; Length 116;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 2 LLSTHLF 8
 DR 95 LLSTHLF 101

RESULT 8
 ID INVO_CANFA
 ID INVO_CANFA
 AC P18174;
 DT 01-NOV-1990 (Rel. 16, Created)
 DT 01-NOV-1990 (Rel. 16, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Involutrin.
 GN INV.
 OS Canis familiaris (Dog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Carnivora; Fissipedia; Canidae; Canis.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=90348475; PubMed=2385171;
 RA Tseng H, Green H;
 RT "The involutrin genes of pig and dog: comparison of their segments of
 repeats with those of prosimians and higher primates.";
 RL Mol. Biol. Evol. 7:291-302(1990).
 CC -1- FUNCTION: Part of the insoluble cornified cell envelope (CE) of
 stratified squamous epithelia.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic. Constituent of the scaffolding
 CC of the cornified envelope.
 CC -1- TISSUE SPECIFICITY: Keratinocytes of epidermis and other
 CC stratified squamous epithelia.
 CC -1- PTM: Substrate of transglutaminase. Specific glutamines or lysines
 are cross-linked to keratins, desmoplakin and to inter involutrin
 CC molecules.

RESULT 7
 FLHD_XENNE

ID	FLHD_XENNE	STANDARD:	PRT;	116 AA.
AC	Q9XF2;			
DT	30-MAY-2000 (Rel. 39, Created)			
DT	30-MAY-2000 (Rel. 39, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Flagellar transcriptional activator flhd.			
GN	FLHD.			
OS	Xenorhabdus nematophilus.			
OC	Xenorhabdus.			
OC	Xenorhabdus.			
OX	NCBI_TaxID=628;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=FL.			
RA	Givaudan A.G., Lanois A.;			
RT	"flhd gene disruptions leads to pleiotropic phenotypes.";			
RL	Submitted (NOV-1988) to the EMBL/Genbank/PDBJ databases.			
CC	-1- FUNCTION: Transcriptional activator. Together with flhc it acts as a compound sigma factor that activates class 2 flagellar genes (By similarity).			
CC	-1- SIMILARITY: BELONGS TO THE FLHD FAMILY.			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/ or send an email to license@isb-sib.ch).			
CC	DR EMBL; AU12838; CAB41407.1; - KW Transcription; Transcription regulation; Sigma factor; CC DNA-directed RNA Polymerase; DNA-binding; Flagellin. CC SQ SEQUENCE 116 AA; 1303 MW; 89DBP#817532828 CRC64;			
QY	2 LLSTHLF 8			
DR	95 LLSTHLF 101			

CC -!- SIMILARITY: BELONGS TO THE INVOLUCRIN FAMILY.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

CC or send an email to license@isb-sib.ch).

EMBL; M3442; AAA30853.1; -

KW PROSITE; PS00795; INVOLCRIN; 1.

DR InterPro; IPR00230; Involutin.

DR Keratinization, Repeat.

SEQUENCE 285 AA; 33384 MW; DCE1BD88B9248BEA CRC64;

QY 32 KQQKEQ 38

Db 198 KQDQEQQ 204

Best Local Similarity 100.0%; Pred. No. 6; 5; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 9

Y687_HAETN STANDARD; PRT; 304 AA.

ID Y687_HAETN STANDARD; PRT; 304 AA.

AC P71356; 16-OCT-2001 (Rel. 40, created)

DT 16-OCT-2001 (Rel. 40, Last sequence update)

DT 16-OCT-2001 (Rel. 41, Last annotation update)

DE Hypothetical transport protein H10687.

GN H10687.

OS Haemophilus influenzae.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Haemophilus; OX NCBI_TAXID=727;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN_RD / KW20 / ATCC 51907;

RX MEDLINE=95350630; PubMed=7542800;

RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F., Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M., Mckenney K., Sutcliffe G., Fitzhugh W., Fields C.A., Gocayne J.D., Scott J.D., Shirley R., Liu L.-I., Goedek A., Kelley J.M., Weidman J.F., Phillips C.A., Spriggs T., Heidelberg J., Cotton M.D., Utterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C., Fine L.D., Fitchman J.L., Fuhrmann J.L., Geoghegan N.S.M., Grahame C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O., Venter J.C., "Whole genome random sequencing and assembly of Haemophilus influenzae Rd.", Science 269:496-512(1995).

RL -!- SUBCELLULAR LOCATION: Integral membrane protein (Probable).

CC -!- SIMILARITY: BELONGS TO THE EAMA TRANSPORTER FAMILY.

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; 032751; AAC22347.1; -

DR InterPro; IPR000620; DUF6.

DR Pfam; PP00892; DUR6; 2.

KW Hypothetical protein; Transport; Transmembrane; Complete proteome.

FT TRANSMEM 9 29

FT TRANSMEM 67 87

FT TRANSMEM 100 120

POTENTIAL.

FT TRANSMEM 131 151

FT TRANSMEM 159 179

FT TRANSMEM 189 209

FT TRANSMEM 222 242

FT TRANSMEM 252 272

FT TRANSMEM 278 298

FT TRANSMEM 304 AA; 33887 MW; CC7095529EB4FB3 CRC64;

QY 7 LFIVLIFI 13

Db 71 LFIVLIFI 77

Best Local Similarity 100.0%; Pred. No. 6; 9; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 10

YK8_YEAST STANDARD; PRT; 337 AA.

ID YK8_YEAST STANDARD; PRT; 337 AA.

AC P6164; 01-JUN-1994 (Rel. 29, Created)

DT 01-JUN-1994 (Rel. 29, Last sequence update)

DT 01-JUN-1994 (Rel. 29, Last annotation update)

DE Hypothetical protein in PRP16-SRP40 intergenic region.

GN YKR088C OR YKR408

OS Saccharomyces cerevisiae (Baker's yeast).

OC Saccharomyces cerevisiae; Ascomycota; Saccharomycotina; Saccharomycetes; Ox

NCBI_TAXID=932;

[1]

RP SEQUENCE FROM N.A.

RX MEDLINE=94263327; PubMed=8203164;

RA Garcia-Cantalejo J., Baldador V., Esteban P.F., Santos M.A., Bou G., Remacha M.A., Revuelta J.P.G., Jimenez A., del Rey F.;

RT "The complete sequence of an 18,002 bp segment of Saccharomyces cerevisiae chromosome XI contains the HBS1, MRP-L20 and PRP16 genes, and six new open reading frames.";

RL Yeast 10:231-245(1994).

CC -!- SUBCELLULAR LOCATION: Integral membrane protein (Potential).

This SWISS-PROT entry is copyright. It is produced through a collaboration

between the Swiss Institute of Bioinformatics and the EMBL outstation -

the European Bioinformatics Institute. There are no restrictions on its

use by non-profit institutions as long as its content is in no way

modified and this statement is not removed. Usage by and for commercial

entities requires a license agreement (See <http://www.isb-sib.ch/announce/>

or send an email to license@isb-sib.ch).

EMBL; 227116; CA81639.1; -

DR EMBL; 228313; CA82167.1; -

DR PIR; S38166; S38166.

DR PIR; S39129; S39129.

DR SGD; S0001796; YKR088C

KW Hypothetical protein; Transmembrane.

FT TRANSMEM 96 115

FT TRANSMEM 138 162

FT TRANSMEM 173 191

FT TRANSMEM 222 246

FT TRANSMEM 253 271

FT TRANSMEM 287 309

FT TRANSMEM 337 AA; 38311 MW; 7EA95DD4E5AF77FE CRC64;

QY 7 LFIVLIFI 13

Db 264 LFIVLIFI 270

Best Local Similarity 100.0%; Pred. No. 7; 5; Mismatches 0; Indels 0; Gaps 0;

Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

RESULT 11

CITC_ECOLI
ID CITC_ECOLI STANDARD; PRT; 352 AA.
AC P77390; 054337; Q9R74;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE [Citrate (pro-3S)-lyase] ligase (EC 6.2.1.22) (citrate lyase
synthetase) (Acetate:SH-citrate lyase ligase).
DE CITC OR B0618.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC NCBI_TAXID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F., Plunkett G., III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of *Escherichia coli* K-12.";
RL Science 277:1453-1474(1997).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshina T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Kimura K., Inada T., Itoh T., Kojihara M., Kanai K., Kasahimoto K.,
RA Mori H., Motomura K., Nakamura Y., Nishio Y., Saito N.,
RA Sampei G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horikuchi T.;
RT "A 718-kb DNA sequence of the *Escherichia coli* K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
RN [4]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RA Ingmer H., Cohen S.N.;
RL Submitted (JUN-1998) to the EMBL/GenBank/DBJ databases.
CC -1- FUNCTION: ACETYLATION OF THE GAMMA SUBUNIT OF CITRATE LYSE.
CC -3'-DIPHOSPHOCORNYMATE-A) OF THE PROSITHEMIC GROUP (2-(5''-PHOSPHORIBOSYL)-
CC -1- CATALYTIC ACTIVITY: ATP + acetate + [citrate (pro-3S)-lyase]
CC (thiol form) = AMP + diphosphate + [citrate (pro-3S)-lyase]
CC (acetyl form).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
use by non-profit institutions as long as its content is in no way
modified and this statement is not removed. Usage by and for commercial
entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
or send an email to license@isb-sib.ch).
CC -----
DR Ecolene; EG13645; cit.
DR InterPro; IPR05216; cit_ly_ligase.
DR InterPro; IPR04821; cyt_tren_rel.
DR InterPro; IPR00182; Gcn5acetyltransfer.
DR Pfam; PF00583; Acetyltransfer; 1.
DR TIGRFAMS; TIGR0014; cit_ly_ligase; 1.
DR TIGRFAMS; TIGR00125; cyt_trn_rel; 1.

KW Ligase; complete proteome.
SQ SEQUENCE 352 AA; 40077 MW; F5894FFCD0F06518 CRC64;
Query Match 8 6%; Score 7; DB 1; Length 352;
Best Local Similarity 100.0%; Pred. No. 7.8;
Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
Oy 4 STHLFY 10
Db 90 STHLIX 96

RESULT 12
ROX1_YEAST STANDARD; PRT; 368 AA.
ID ROX1_YEAST
AC P25042;
DT 01-MAY-1992 (Rel. 22, Created)
DT 01-MAY-1992 (Rel. 22, last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)
DE ROXI repressor (hypoxic function repressor) (Heme-dependent repression factor).
GN ROXI OR YPR065W OR IPK949.20.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycotina; Saccharomycetes;
OC Saccharomycetidae; Saccharomyctaceae; Saccharomyces.
OX NCBI_TAXID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=94019382; PubMed=8413209;
RA Balasubramanian B., Lowy C.V., Zitomer R.S.;
RT "The ROXI repressor of the *Saccharomyces cerevisiae* hypoxic genes is
a specific DNA-binding protein with a high-mobility-group motif.";
RL Mol. Cell. Biol. 13:6071-6078(1993).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=9731371; PubMed=9159875;
RA Bussey H., Stoeckli R.K., Ahmed A., Albermann K., Allen E., Ansorge W.,
RA Araullo R., Aparricio A., Barrell B.G., Badcock K., Benes V.,
RA Botstein D., Bowman S., Bruckner M., Carpenter J., Cherry J.M.,
RA Chung E., Churcher C.M., Coster F., Davis K., Davis R.W.,
RA Dietrich F.S., Delius H., DiPaolo T., Dubois E., Duesterhoeft A.,
RA Duncan M., Floeth M., Fortin N., Friesen J.D., Fritz C., Goffeau A.,
RA Hall J., Hebling U., Heumann K., Hilbert H., Hillier D.,
RA Hunnicke-Smith S., Hyman R., Johnston M., Kilian S., Kleine K.,
RA Komp C., Kurdi O., Lashkari D., Lew H., Lin A., Lin D., Louis E.J.,
RA Marchal R., Messenguy F., Mewes H.-W., Mirtzpati S., Moestl D.,
RA Mueller-Auer S., Namath A., Nentwich U., Offen P., Pearson D.,
RA Pettel F.X., Pohl T.M., Purnelle D., Schafer M., Schafre M.,
RA Scherrenb. S., Schramm S., Schroeder M., Sdica A.M., Tettelin H.,
RA Teruel L.A., Ushinsky S., Vierendeels F., Vissers S., Voss H.,
RA Walsh S.V., Wambutt R., Wang Y., Wedler E., Wedler H., Winnett E.,
RA Zhang W.W., Zolnier A., Vo D.H., Hani J.;
RA "The nucleotide sequence of *Saccharomyces cerevisiae* chromosome XVI.";
RN Nature 387:103-105(1997).
RN [3]
RP CHARACTERIZATION.
RX MEDLINE=96174644; PubMed=8600445;
RA di Flumeri C., Liston P., Acheson N.H., Keng T.;
RA "The HMG domain of the ROXI protein mediates repression of HEM13
through overlapping DNA binding and oligomerization functions.";
RL Nucleic Acids Res. 24:808-815(1996).
CC -1- FUNCTION: TRANSCRIPTION FACTOR THAT REPRESSES THE EXPRESSION OF
CC HEM13, COX5B, ANBL, CYC7 OR AAC3 (HYPOXIC FUNCTION). BINDS TO THE
CC DNA SEQUENCE 5'-RRPAAAGAC-3'.
CC -1- SUBCELLULAR LOCATION: Nuclear.
CC -1- INDUCTION: BY HEME.
CC -1- SIMILARITY: CONTAINS 1 HMG BOX.

This SWISS-PROT entry is copyright. It is produced through a collaboration
between the Swiss Institute of Bioinformatics and the EMBL outstation -
use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC
 DR EMBL; XE0458; CAA2991_1; --;
 DR EMBL; ZA9219; CAA9182_1; --;
 DR PIR; S1'015; S17015.
 DR HSSP; Q5066; IHRX.
 DR TRANSFAC; T01286; --.
 DR SGD; S006269; R0X1.
 DR InterPro; IPR00910; HMG_12_box.
 DR Pfam; PF00505; HMG_box; 1.
 DR SMART; SM0098; HMG; 1.
 KW Transcription regulation; Repressor; DNA-binding; Nuclear protein.
 FT DNA_BIND 14 83 GLN-RICH
 DOMAIN 102 123 CRC64;
 SEQUENCE 368 AA: 41838 MW: 3B2742D7DEB3DBD CRC64;
 QY Query Match 8 6%; Score 7; DB 1; Length 368;
 Best Local Similarity 100.0%; Pred. No. 8 1; Mismatches 7; Indels 0; Gaps 0;
 Db 105 QQREQQ 112

RESULT 13
 ALU_HUMAN STANDARD PRT; 591 AA.
 ID ALU_HUMAN STANDARD PRT; 591 AA.
 AC P39188;
 DT 01-FEB-1995 (Rel. 31; Created)
 DT 01-FEB-1995 (Rel. 31; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; Last annotation update)
 DE Alu subfamily J sequence contamination warning entry.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TAXID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-95021758; PubMed=7935834;
 RA Claverie J.-M., Makalowski W.;
 RT "Alu alert.";
 RL Nature 371:752-752(1994).
 RN [2]
 RP CONCEPT
 RX MEDLINE-92241891; PubMed=1572661;
 RA Claverie J.-M.;
 RT "Identifying coding exons by similarity search: alu-derived and other
 potentially misleading protein sequences.";
 RL Genomics 12:838-841(1992).
 RN [3]
 RP ALU FAMILIES CLASSIFICATION.
 RX MEDLINE-8833009; PubMed=3138422;
 RA Quentin Y.;
 RT "The Alu family developed through successive waves of fixation
 closely connected with primate lineage history.";
 RL J. Mol. Evol. 27:194-202(1988).
 RN [4]
 RP ALU FAMILIES CLASSIFICATION
 RX MEDLINE-91178815; PubMed=1706781;
 RA Junka J., Milosav-Jevic A.;
 RT "Reconstruction and analysis of human Alu genes.";
 RL J. Mol. Evol. 32:105-121(1991).
 CC -1- MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE
 THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
 CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
 FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
 REPEATS.

CC -1- MISCELLANEOUS: ISOLATED 'X' INDICATES THE PRESENCE OF A STOP
 CODON, 'XXX' IS USED TO SEPARATE THE VARIOUS TRANSLATION PHASES.

CC -1- CAUTION: THIS ALU ENTRY IS PROVIDED IN ORDER TO AVOID THE FURTHER
 CC POLLUTION OF PROTEIN SEQUENCE DATABASES WITH ALU-DERIVED AMINO
 CC ACID SEQUENCES.
 CC -1- CAUTION: ALU REPETITIVE SEQUENCES ARE INTERSPERSED IN HUMAN AND
 CC PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
 CC ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
 CC ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAS
 CC LIGATED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
 CC ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
 CC GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS, CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
 CC OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
 CC DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
 CC THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
 CC BEING REPORTED.
 CC -1- CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
 CC WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
 CC PART OF ALU REPEAT MAY HAVE BEEN ARTIFICIALLY INCLUDED IN THE
 CC CODING NUCLEOTIDE SEQUENCE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement. (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC
 CC DR EMBL; U14567; -; NOT_ANNOTATED_CDS.
 CC KW Hypothetical protein.
 CC FT DOMAIN 1 96 FRAME-1.
 CC FT DOMAIN 100 195 FRAME-2.
 CC FT DOMAIN 199 294 FRAME-3.
 CC FT DOMAIN 298 393 FRAME-4.
 CC FT DOMAIN 397 492 FRAME-5.
 CC FT DOMAIN 496 591 FRAME-6.
 CC SQ SEQUENCE 591 AA: 63790 MW: 665D395735519D95 CRC64;
 QY Query Match 8 6%; Score 7; DB 1; Length 591;
 Best Local Similarity 100.0%; Pred. No. 12; Mismatches 7; Indels 0; Gaps 0;
 Db 19 SLGDRR 25
 11111
 85 SLGDRR 91

RESULT 14
 ALU4_HUMAN STANDARD PRT; 603 AA.
 ID ALU4_HUMAN STANDARD PRT; 603 AA.
 AC P39191;
 DT 01-FEB-1995 (Rel. 31; Created)
 DT 01-FEB-1995 (Rel. 31; Last sequence update)
 DT 16-OCT-2001 (Rel. 40; last annotation update)
 DE Alu subfamily SB2 sequence contamination warning entry.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
 NCBI_TAXID=9606;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE-95021758; PubMed=7935834;
 RA Claverie J.-M., Makalowski W.;
 RT "Alu alert.";
 RL Nature 371:752-752(1994).
 CC [2]
 RP CONCEPT
 RX MEDLINE-92241891; PubMed=1572661;
 RA Claverie J.-M.;
 RT "Identifying coding exons by similarity search: alu-derived and other
 potentially misleading protein sequences.";

RL Genomics 12:638-841(1992).
 RN [3] ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=88333009; PubMed=3138422;
 RA Quentin Y.;
 "The Alu family developed through successive waves of fixation
 closely connected with primate lineage history.";
 J. Mol. Evol. 27:194-202(1988).
 RL [4] ALU FAMILIES CLASSIFICATION.
 RX MEDLINE=91178015; PubMed=1706781;
 RA Jurija J., Milosavljevic A.;
 "Reconstruction and analysis of human Alu genes.";
 J. Mol. Evol. 32:105-121(1991).
 RL [-1] MISCELLANEOUS: VARIOUS ANALYSES (SEE REF. 3 AND REF. 4) INDICATE
 THAT ALU REPEATS FALL INTO 8 SUBFAMILIES. THEREFORE, 8 ALU WARNING
 CONSENSUS SEQUENCES HAVE BEEN CONSTITUTED THAT CONTAIN ALL SIX
 FRAMES CONCEPTUAL TRANSLATIONS OF EACH OF THESE CLASSES OF ALU
 REPEATS.
 CC [-1] CAUTION: ALU REPETITIVE SEQUENCES ARE INTERPERSSED IN HUMAN AND
 PRIMATE GENOMES WITH AN AVERAGE SPACING OF 4 KB. SOME OF THEM ARE
 ACTIVELY TRANSCRIBED BY POL III. NORMAL TRANSCRIPTS MAY CONTAIN
 ALU-DERIVED SEQUENCES IN 5' OR 3' UNTRANSLATED REGIONS. HOWEVER,
 CC CDNA LIBRARIES ALSO CONTAIN PARTIAL AND/OR REARRANGED CDNAs.
 CC LIGHTED WITH ALU-DERIVED SEQUENCE IN ANY ORIENTATION. ALTHOUGH ALU
 ELEMENTS (ESPECIALLY SITUATED ON THE COMPLEMENTARY STRAND) HAVE A
 GREAT POTENTIAL TO CREATE ADDITIONAL/ALTERNATIVE EXONS,
 CONSIDERATION SHOULD BE GIVEN TO THE POSSIBILITY THAT THE PRESENCE
 OF AN ALU IN AN OPEN READING FRAME MAY HAVE RESULTED FROM A
 CC CLONING ARTIFACT OR MAY BE DUE TO MISINTERPRETATION OF SEQUENCING
 DATA. THIS POINT HAS BEEN OVERLOOKED ON SEVERAL OCCASIONS, WITH
 THE CONSEQUENCE OF ERRONEOUS ALU-DERIVED AMINO ACID SEQUENCES
 BEING REPORTED.
 CC [-1] CAUTION: ANY SIGNIFICANT SIMILARITY OF A PUTATIVE PROTEIN SEQUENCE
 WITH AN ALU-TRANSLATED ENTRY MUST BE TAKEN AS A WARNING THAT A
 PART OF ALU REPEAT MAY HAVE BEEN ARTIFICIALLY INCLUDED IN THE
 CODING NUCLEOTIDE SEQUENCE.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 DR EMBL; U14570; --; NOT_ANNOTATED_CDS.
 KW Hypothetical protein.
 FT DOMAIN 1 98 FRAME-1.
 FT DOMAIN 102 199 FRAME-2.
 FT DOMAIN 202 300 FRAME-3.
 FT DOMAIN 304 401 FRAME-4.
 FT DOMAIN 405 502 FRAME-5.
 FT DOMAIN 506 603 FRAME-6.
 SQ SEQUENCE 603 AA; 65272 MN; BRAD0046BEA114 CRC64;
 CC Query Match 8.6%; Score 7; DB 1; Length 603;
 CC Best Local Similarity 100.0%; Pred. No. 12;
 CC Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 20 LGDRRL 26
 Db 8B LGDRRL 94
 RESULT 15
 SYL_BACSU ID STANDARD; PRT; 804 AA.

AC P36430; 034455;
 DT 01-JUN-1994 (Rel. 29, Created)
 DT 15-JUL-1998 (Rel. 36, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Leucyl-tRNA synthetase (EC 6.1.1.4) (Leucine-tRNA ligase) (LeuRS).
 GN LETS.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 RX NCBI_TAXID=1433;
 RN [1] SEQUENCE FROM N.A.
 RP MEDLINE=92283747; PubMed=1317842;
 RA Vander Horn P.B., Zahler S.A.;
 RT "Cloning and nucleotide sequence of the leucyl-tRNA synthetase gene
 in the 200 kb rrnB-dnaB region.";
 RL J. Bacteriol. 174:3928-3935(1992).
 RN [2] SEQUENCE FROM N.A.
 RP STRAIN=168;
 RN MEDLINE=96048467; PubMed=9387221;
 RC STRAIN=168;
 RX Lapidus A., Galleron N., Sorokin A., Ehrlich S.D.;
 RA "Sequencing and functional annotation of the *Bacillus subtilis* genes
 in the 200 kb rrnB-dnaB region.";
 RT Microbiology 143:3431-3441(1997).
 RL [3] SEQUENCE FROM N.A.
 RN MEDLINE=98044333; PubMed=9384377;
 RA Kunst F., Ogakawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azereedo V., Berti M.G., Bessieres P., Bolotin A., Borchart S.,
 RA Borrius R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brodinlet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Coelani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
 RA Denizot F., Devine K.M., Dusterhoff A., Ehrlich S.D., Emmerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrall E., Fouger D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghini S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Gubesspi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Hollsappel S., Hosono S., Hullio M.F., Itaya M., Jones L.,
 RA Joris B., Karimata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetteler P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidozzi A., Hardinno C., Laufer J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Maveli C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parrot V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Presecan E., Pujic P., Purcell B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadale Y.,
 RA Satoh T., Scanian E., Schleicher R., Schroeter R., Scoffone F.,
 RA Soguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
 RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terpetra P., Tognoni A.,
 RA Totsuka V., Uchiyama S., Vandenberg M., Vanner F., Vassarotti A.,
 RA Viari A., Wambutt R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winter P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yosikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram positive bacterium *Bacillus subtilis*";
 RT Nature 390:249-256(1997).
 CC [-1] CATALYTIC ACTIVITY: ATP + L-leucine + tRNA(Ieu) = AMP +
 CC diphosphate + L-leucyl-tRNA(Ieu).
 CC [-1] SUBCELLULAR LOCATION: Cytoplasmic.
 CC [-1] SIMILARITY: BELONGS TO CLASS I AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 between the Swiss Institute of Bioinformatics and the EMBL outstation -
 the European Bioinformatics Institute. There are no restrictions on its
 use by non-profit institutions as long as its content is in no way
 modified and this statement is not removed. Usage by and for commercial
 entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 or send an email to license@isb-sib.ch).
 CC
 EMBL; MB8581; AAA22571.1; .

DR EMBL; AF008220; AAC00259.1; -.
 DR EMBL; Z93119; CAB15010.1; -.
 DR PIR; A41882; A41882.
 DR Sublist; BG10675; leus.
 DR InterPro; IPR02307; Leu-tRNA-syntia.
 DR InterPro; IPR02300; tRNA-synt_1a.
 DR InterPro; IPR01412; tRNA-synt_I.
 DR Pfam; PF00133; tRNA-synt_1; 1.
 DR PRINTS; PR0085; TRNASYNTHEL0.
 DR TIGRFAMS; TIGR00316; leus_bct; 1.
 DR PROSITE; PS00178; AA_tRNA_LIGASE_I; 1.
 DR PROSITE; PS00178; AA_tRNA_Synthetase; Protein biosynthesis; Ligase; AMP-binding;
 KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; AMP-binding;
 KW Complete proteome.
 FT SITE 40 51 "HIGH" REGION.
 FT SITE 576 580 "KMSKS" REGION.
 FT BINDING 579 579 ATP (BY SIMILARITY).
 FT CONFLICT 186 186 P -> L (IN REF. 1).
 FT CONFLICT 195 195 T -> N (IN REF. 1).
 FT CONFLICT 247 281 RPDILFGAYTVLAPHEAHVENTITAEOREADEVAY -> DQ
 FT SEQUENCE 804 AA; 91542 MW; 306FD5A9BFEC47E CRC64;
 REF. 1).
 FT SEQUENCE 804 AA; 91542 MW; 306FD5A9BFEC47E CRC64;
 Query Match 8.6%; Score 7; DB 1; Length 804;
 Best Local Similarity 100.0%; Pred. No. 16;
 Matches 7; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 Qy 56 TGKKGR 62
 Db 479 TGKKGR 485

Search completed: May 29, 2003, 15:31:09
 Job time : 23 secs